DOCUMENT RESUME

ED 174 201 IR 007 475

AUTHOR Wellisch, Hans H.
TITLE The Exchange of Bibliographic Data in Non-Roman Scripts.
PUB DATE Apr 78
NOTE 16p.; Paper presented at the International Symposium on Bibliographic Exchange Formats (Taormina, Sicily, April 27-29, 1978)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cataloging; International Programs; *Problems; *Romanization

ABSTRACT

Documents in non-Roman scripts now constitute a sizeable part of world production, and their bibliographic control through Romanization is beset by many problems. Among those are the impossibility to simultaneously satisfy certain functional requirements, the multiplicity of schemes and their inconsistent use, and the susceptibility of the method to errors. New and more effective methods must therefore supplement Romanization. Efforts should be concentrated on revising or creating transliteration standards based on official national schemes where available, and with no regard to phonetic or phonemic features; and designing machine-readable conversion codes for the 10 major scripts to be used exclusively for exchange purposes. International documentation services should use only international conversion standards, while libraries and retrospective bibliographies should use local or national conversion schemes best suited to the sociolinguistic habits of their users, supplemented by listings of documents in non-Roman scripts. Indexing and abstracting services should Romanize only names of authors and titles of source periodicals but not titles of individual articles, and bibliographic data in non-Roman scripts should be provided primarily by the respective national bibliographies, both in camera-ready copy and, if possible, in machine-readable form. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED174201

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE EXCHANGE OF BIBLIOGRAPHIC DATA IN NON-ROMAN SCRIPTS

by

Hans H. Wellisch

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Hans H. Wellisch

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

IR000747

THE EXCHANGE OF BIBLIOGRAPHIC DATA IN NON-ROMAN SCRIPTS

Hans H. Wellisch*

ABSTRACT

It is shown that documents in non-Roman scripts constitute now a sizeable part of world production and that their bibliographic control through Romanization is beset by many problems. Among these are the impossibility to simultaneously satisfy certain functional requirements (e.g. reversibility, simple graphics, uniqueness of graphic representation), the multiplicity of schemes and their inconsistent use, the susceptibility of the method to errors, and others. When only Romanization is used for the bibliographic control of non-Roman documents, no real exchange of data is taking place, because it is a one-directional process. In most instances, unambiguous identification of documents is not possible. New and more effective methods must therefore supplement Romanization.

It is proposed to abandon efforts to create "international" Romanization standards suitable for all purposes and pleasing all users because this is impossible to achieve. Rather, efforts should be concentrated on (a) revising or creating transliteration standards mainly based on official national schemes where available, and with no regard to phonetic or phonemic features; (b) designing machine-readable conversion codes for the ten major scripts, to be used exclusively for exchange purposes.

International documentation services should use only international conversion standards. Libraries and retrospective bibliographies, including national bibliographies, should use local or national conversion schemes best suited to the sociolinguistic habits of their users, supplemented by listings of documents in non-Roman scripts (in the original). Indexing and abstracting services should Romanize only names of authors and titles of source periodicals but not titles of individual articles; from time to time, they should publish lists of sources both in Romanized form and in the original script to allow unambiguous identification.

Bibliographic data in non-Roman scripts should be provided primarily by the respective national bibliographies, both in camera-ready copy and, if possible, in machine-readable form. Separate data bases would thus be established for each of the major scripts.

These proposals recognize that (a) Romanization is necessary for Western bibliographic control tools; (b) existing Romanization systems cannot easily be changed or abandoned for both economical and technical reasons; (c) they must be supplemented by displays of bibliographic data in their original form to make unambiguous identification possible; (d) that efficient and effective exchange of bibliographic data is best being achieved by the use of machine-readable switching codes.

* Associate Professor, College of Library and Information Services, University of Maryland, College Park, MD 20742.

THE EXCHANGE OF BIBLIOGRAPHIC DATA IN NON-ROMAN SCRIPTS

Effective and efficient exchange of bibliographic data on a national and international scale must be based on standards that assure compatibility and mutual intelligibility. Few people would question the validity of this statement, but there is much less agreement on the methods and systems which ought to be used to accomplish this aim. Moreover, this task would be difficult enough even if all the world's written records were produced in the same script, but, as everybody knows, such is not the case. The problem posed by documents written in different scripts is, however, generally thought to be of relatively minor importance on a global scale, and it is also assumed as axiomatic that the conversion of non-Roman scripts into Roman characters -- Romanization -- assures exact bibliographic control and makes the unambiguous exchange of bibliographic data possible. Neither of these assumptions is correct.

The number of documents in non-Roman scripts

We shall first consider the quantitative aspect. Available data for the world's annual book production (see Table 1) show that more than 160,000 titles or almost 30% of the total is published in non-Roman scripts. Almost exactly half of this quantity is in Cyrillic script, followed by titles in Japanese and Chinese, Devanagari, Arabic and other scripts in descending order. The current world production of articles in science (in the widest sense) and technology is conservatively estimated at about 2,500,000; this figure is based on a total of presently existing scholarly periodicals numbering about 35,000, each of which produces an average of 70 articles per year. No recent figures on the distribution of articles by language and script are available, but assuming (again very conservatively) that only 15% of the world total is printed in non-Roman scripts (primarily in Cyrillic and in Japanese which accounted for 11% and 4% respectively according to an estimate made in 1962, and allowing for a mere 1% for contributions in other scripts which have become prominent since) the number of articles presently published in non-Roman scripts would be about 375,000 per year.[1] The total number of all printed publications in non-Roman scripts can thus be estimated to be at least 500,000 -- certainly not an insignificant quantity, even though a certain part of these publications is of purely local interest[2] and thus not necessarily subject to Universal Bibliographic Control. But if only two thirds of all documents in non-Roman scripts are recorded in bibliographic control tools (library catalogs, bibliographies, and abstracting and indexing services) outside their country of origin, this would mean that there are at least 330,000 items per year to be handled, and their bibliographic data have to be exchanged.

Romanization

The rendering of non-Roman scripts by means of Romanization, either by representing each foreign character by one or more corresponding Roman

letters (transliteration) or by an approximation of the sound of words (transcription) is a time-honored practice. It is indeed a necessity in translation, in the writing of works concerned with the history, geography, and culture of countries in which a non-Roman script is being used, and when news from those countries have to be reported; it is also indispensable in cartography. But in the realm of bibliographic control it can fulfil only an ancillary role, and it should certainly not serve as the sole means for the effective exchange of bibliographic data, at least not in its traditional form.

The shortcomings and pitfalls of Romanization have recently been pointed out by several writers[3] and the underlying reasons for its inadequacy as a bibliographic control tool have been explored in detail elsewhere.[4] In this context, these can only be summarized briefly.

Conflicting requirements

There are at least six requirements which Romanization systems for bibliographic control purposes have variously been expected to fulfil, namely:

(a) general applicability (the system must be able to render any word or name in the source script in Roman letters);
(b) uniqueness of graphemic representation (each letter or character in the source script must be represented unambiguously);
(c) simplicity of graphemic representation (use of basic letters only with no or few diacritical marks);
(d) reversibility (the characters of the non-Roman source script must be unambiguously restorable by simply reversing the transformation process);
(e) traditionality (Romanization of names that have a long-standing tradition in a language community, e.g. Biblical names, must be preserved);
(f) pronounceability (the concatenation of Roman letters produced by Romanization must be pronounceable, especially for names of persons and places).

All of these requirements except the first one (which is almost trivial) clash with each other, even if for the more restricted purpose of exchange of bibliographic data the last two requirements can be disregarded as not relevant. For any non-Roman source script, only one or at best two of the requirements can be fulfilled simultaneously, while some or all of the others must be sacrificed. For certain scripts, some requirements are even impossible to fulfil, no matter which system is being used; for example, no Romanization of Chinese or Japanese can be reversible, and any alphabet larger than the Roman one must necessarily be Romanized by using a number of digraphs or trigraphs (which may be ambiguous) or by diacritical marks, or both.

Multiplicity of Romanization schemes

Romanization as presently practiced by the world's large national libraries, by national bibliographies and by national and international

agencies and services is not applied consistently and exchangeably because a large number of various systems are in use. Some organizations use ISO transliteration standards, some cling to systems with a long bibliographic tradition such as the schemes of the Library of Congress or those of the Preussische Instruktionen, some use the national standards of their respective countries, and some others use various locally devised adaptations and mixtures of these systems.[5] None of these systems is unambiguously reversible.

Although users of a particular brand of Romanized bibliographic records may become accustomed to the idiosyncrasies of the system, probably no one in need of information contained in non-Roman documents can rely exclusively on one bibliographic source only. The moment a user turns to several national bibliographies he may find, for example, the Russian name Юшкевич Romanized in three or four different ways, filed alphabetically in as many different places; if he then adds references from an international bibliographic service such as INIS he may be confronted with yet another transcription compatible neither with any of the national systems nor with the international systems of ISO. Occasionally, a bibliography will contain the name of an author in two different forms, depending on whether his works are cited in the (transliterated) original or in translation.

Susceptibility to errors

The process of Romanization is also exceedingly prone to errors because of the complexity of the transformations that have to be performed. All systems presently in use presuppose a knowledge of the source language and its orthographic conventions, and they contain a sometimes quite large number of special provisions and exceptions to the general rules. In many instances, the rules of a Romanization system are set aside entirely when considerations of convention and tradition seem to be more important, as in the case of Biblical names, names from classical antiquity, and in other cases. When titles are being Romanized as a string of letters forming Romanized words, the result cannot be verified in dictionaries because such reference tools do not exist (except for Japanese, in which case, however, a thorough knowledge of the language and its writing system is also needed for verification, and mere "look-up" by a Westerner is almost totally useless).

For these and other reasons which cannot be dealt with here for lack of space, Romanization is not a reliable control tool for the unambiguous identification of documents, and it serves thus neither librarians, bibliographers and indexers, nor the end users of a bibliographic system as well as is generally assumed.

It has, of course, time and again be stated that all or most of the difficulties would disappear if only all concerned would agree on one single internationally accepted standard system (such as the ISO transliteration standards). Yet this is unlikely to happen in the foreseeable future, for a variety of technical, economic and sociolinguistic reasons. There is no need to dwell on the technical and economic dif-

ficulties which are obvious, but perhaps a hint about the sociolinguistic problems can be given. Among these are long-standing orthographic conventions such as the presence or absence of diacritical marks in the writing system of certain language communities; the completely different pronunciation of several Roman letters (mainly c, g, h, j, s, v, and x) in various languages; and the psycholinguistic phenomenon that the reader of a Romanized name or word feels compelled to verbalize or at least subvocalize it according to the phonetics of his writing system.

Bibliographic colonialism

Finally, Romanization (even if it could be performed according to one internationally accepted system) can never be a tool of exchange. The concept of "exchange" entails the notion of "giving and taking reciprocally" but Romanization is by definition a one-way operation only, based on the idea that the written records of mankind can be brought under control only if they are recast in the mold of Roman letters, all other modes of writing seemingly being either unsuitable or inferior or both. Although nobody would probably say so aloud, there seems to be an unconscious hope on the part of fervent Romanizers that it might ultimately be possible to persuade Russians, Japanese, Arabs, and many other peoples to abandon their inscrutable writing systems and to convert them to the one and only redeeming Roman alphabet -- a kind of conversion of the heathens on the graphemic level. What has been said and also done is that those people who continue to write in their various indigenous scripts must all submit to Romanization if they wish that their books and articles are brought to the attention of the rest of the world. Not only that, but they are told that they ought to perform the transformation of their writing systems into a foreign and unsuitable mold all by themselves, preferably in their own national bibliographies, to please Western-dominated bibliographic control. Far from being an exchange, this can only be called "bibliographic colonialism", perhaps the last vestige of the domination of European powers over peoples in Asia and Africa. In fact, the other side of the coin, namely the problem of rendering Western names written in Roman letters into different writing systems by Cyrillization, Japanization, Arabization, and so on, is almost never considered; nor is it realized that all these conversions of scripts can only be performed by transcription, trying to approximate the sound of words, and never by transliteration, the principle so often presumed to be the foundation of truly "scientific" and fully reversible Romanization. Cyrillization or any other transformation of Roman script into a different one is never reversible, because each language written in Roman letters has different sound values for many letters and letter combinations. Further complications arise when one non-Roman script is converted into another, and the result then Romanized. For example, a Japanese name transcribed into Russian and then re-transliterated into Roman will not be recognizable or retrievable.

Alternatives to Romanization in bibliographic control

Past attempts to create "international" systems of script conversion that would do everything for every purpose have ended in dismal failure and

have not produced anything that would be useful for either Universal Bibliographic Control or for the effective international exchange of bibliographic data as an integral part of UBC. No more efforts in this direction should therefore be made. What is needed is a radically new departure. Not utopian pipedreams of making every library, large or small, and every bibliographic service, general or specialized, conform to exactly identical "universal" schemes based exclusively on the Roman alphabet, but a gradual evolvement of nationally valid bibliographic control methods that are responsive to the sociolinguistic habits of users in various language communities with a common writing system, and linked both to other writing systems and to each other through machine-readable switching codes -- the only kind of exchange mechanism on which international agreement could possibly be reached.

Although at first it might sound paradoxical and quite contrary to the idea of UBC and exchange of bibliographic data across national, linguistic and graphemic boundaries, under the system envisaged here each country or language community should continue to employ the Romanization system (or other conversion method) best suited to the traditions and sociolinguistic habits of the users of its libraries and their bibliographic tools. None of the Romanization schemes presently used in large libraries, national bibliographies or in abstracting and indexing services can really claim to be "scientific"; all are mixtures of transcription and transliteration, and few if any are compatible with each other. But if they work somehow they might as well be retained as an aid to library technicians and as a crutch for those who cannot read a non-Roman script yet wish, for some reasons, to know what books in that script have been written and by whom. Any purportedly "new", "international" or "scientific" system would also be deficient in one respect or another, it would not be accepted by certain classes of users,and would only create further confusion, since older entries in existing catalogs and bibliographies could normally not be made to conform with the "new" system.

Three conditions should, however, be fulfilled whenever Romanization is being employed in catalogs or bibliographies, so as to make its use easier and more meaningful:

(a) It should be clearly stated which system of Romanization is being used, including local modifications or exceptions; the key to the system should be displayed prominently to enable users to understand how names and words have been Romanized, and (if such is possible) to enable them to reverse the Romanization to the original form;
(b) the titles of works in non-Roman scripts should not only be Romanized but also translated into the dominant language of the users;
(c) the language of the work should be indicated in a note.

Since the use of different Romanization systems cannot assure full and unambiguous bibliographic control over documents written in non-Roman scripts, much less effective exchange of bibliographic data, it is necessary to estab-

lish alternative and parallel means of control. The organizational, economic, and technological aspects of such alternative bibliographic control tools have been discussed in detail elsewhere[6] and are here summarized only briefly.

Library catalogs and bibliographies

In library catalogs and in most printed bibliographies, Romanized entries should be accompanied by parallel and complementary listings of documents in their original script to which cross references from the Romanized form are to be made. This is not an altogether revolutionary or new idea: many libraries in Europe and in the Soviet Union as well as most Asian libraries are already following this practice by filing entries for different scripts in separate catalogs. In national bibliographies or subject bibliographies comprising entries for works in different scripts, non-Roman entries should be listed after all Roman script entries (or before those if the non-Roman script is the national one), also with suitable cross references from one section to others. The national bibliography of Sri Lanka, printed in 3 separate sections (Sinhala, Tamil, and Roman-script entries) is a good example for such an arrangement.

As more and more library catalogs are being generated from machine-readable data bases, this implies that separate data bases would have to be established for each major non-Roman script, whereas the minor scripts (i.e. those which produce only a relatively small number of books) could probably be handled more economically by conventional methods, primarily photomechanical reproduction from camera-ready copy. In either case, the basic entries in the original script should be provided by the respective national bibliographies or national libraries.

If, for some reason, it is desirable to merge the Roman and non-Roman entries in a printed catalog, the example of the Hebrew part of the New York Public Library's author and title catalog shows that this can be done, utilizing modern computer-controlled phototypesetting.[7]

Abstracting and indexing services

The purposes of abstracting and indexing services as well as the expectations of their users are sufficiently different from those of library catalogs and retrospective bibliographies to need other solutions of the problem of making documents (mainly articles in periodicals and conference proceedings as well as research reports, patents and standards) in non-Roman scripts accessible and intelligible. The main concerns of users are in this case timeliness (which implies a minimum of delay in publication), precise identification of the source, and an abstract that allows the reader to make a decision on the usefulness of an item. To take the last issue first, a document can be useful even if it has not been read in its entirety (the abstract alone may supply sufficient information) and it may not have to be read in the original but a full or even a partial translation will serve. This means that

users may not necessarily be concerned with the exact wording of a title but are more interested to know the language and the script of the item so as to find a suitable translator. This is substantially different from a library catalog whose users generally want to peruse a book themselves and are less often satisfied with the mere information that such and such a book exists. The emphasis on speed of publication will in most instances also preclude laborious and time-consuming typographical display of the original script, and in practically all situations where a computer terminal is used (either on-line or off-line) such a display can presently not be provided (although it would be feasible using available technology).

The names of authors or sponsoring organizations should be Romanized according to the same principles that govern entries in library catalogs. The wholesale Romanization of titles, however, is an exercise in futility and should in all instances be replaced by a translation into the working language of an abstracting or indexing service, followed by an indication of the language. Several large abstracting services are already using this style, and others should be encouraged to follow suit. Title translations can often be taken directly from the original document; if such is not the case, the translation of a title is generally neither more difficult nor more time-consuming than its Romanization and it needs the same kind of expertise, namely a knowledge of the source language, but it should be intellectually more rewarding to persons performing it than a mindless transformation of one alphabet into another.

The exact identification of the source of an article is independent of script as far as pagination is concerned,[8] but in the rendering of titles of periodicals the problem of Romanization is again encountered. It is, however, somewhat easier to handle than the Romanization of authors' names, because Romanized titles of serials are now to a large extent standardized in the ISDS system through the International Centre for the Registration of Serial Publications in Paris. These standard titles in Romanized form serve as a benchmark against which divergent local or national Romanizations can be checked. This system should be supplemented by periodically issued lists of serials in non-Roman scripts in which both the original title and its official Romanization in the ISDS system would be shown. The Centre in Paris would no doubt be able to draw on the resources of other ISDS centers for this purpose, primarily those in the USSR, in Japan, and in Southeast Asia. Most abstracting and indexing services also publish from time to time lists of periodicals scanned, and it would seem to be entirely feasible to display the titles of periodicals written in one of the non-Roman scripts in the original alongside their Romanized forms, so as to make unambiguous identification of the source possible. This need not be a typographical problem either, because a title in any non-Roman script can be taken from the cover or masthead of a periodical and can then be reproduced photomechanically.

Exchange of bibliographic data in non-Roman scripts

To accomplish effective exchange of bibliographic data in a form that is acceptable to a language community with another writing system,

these data must be linked through a switching mechanism that makes them mutually compatible. The creation of such switching tools in the form of internationally accepted machine-readable codes is a task on which most of the work will yet have to be done.

While it has been practically impossible to reach agreement on a Romanization system that would be acceptable even for the limited purpose of bibliographic data exchange among different language communities that are already committed to long-standing orthographic and bibliographic traditions, there is no reason why it should not be possible to design purely machine-readable conversion codes for each of the ten or so major scripts. In such a machine-readable code it does not matter in the least whether a character of the source script is represented by one, two or more characters in the (Roman) target script, or whether such representation is accomplished by the assignment of numbers (as is likely to be the case for the rendering of Chinese characters), as long as these codes are (a) representing only written characters without any regard to phonological elements; (b) are unambiguous and reversible; and (c) are not intended for eye-legible printout.

Once a certain Romanization system has become the standard for nationally produced bibliographic records, it will in most instances be possible to perform the transformation from a machine-held switching code by algorithm and with no or relatively little post-editing by human beings (depending on the source language). An exception would be the retranscription of transliterated names for which obviously no automatic procedure can be devised. The name Baudelaire, for example, when transcribed in a Russian translation and then transliterated back letter for letter into Roman becomes Bodler; the same process applied to a Greek translation would yield Mpontelair; and for Hebrew it might be Bwdlyr. Only human knowledge of context can restore the original spelling of a name.

Such a system will, of course, not be reversible because every national Romanization in use today contains elements of phonemic transcription not present in a machine-readable switching code. The various Romanizations will also be largely incompatible with each other, as they are today. But this will be of little concern as long as entries in the original script are also provided and properly linked to their Romanized, Cyrillized or whatever other converted form, and if, in turn, the original non-Roman forms can be restored and expressed unambiguously through a machine-readable code.

The tasks ahead

The proposals presented here in broadest outline only will no doubt have to be elaborated in detail. It is hoped that they will prove to be feasible, since they will not disrupt existing bibliographic data structures, will not demand the expenditure of large sums to discard old systems and to switch to entirely new (and unproven) ones, and are making use of existing technology at a relatively low level of sophistication.

Consequently, international organizations, foremost among them UNISIST and ISO, should concentrate their efforts aimed at the integration of bibliographic records written in non-Roman scripts into international control systems on the following tasks.

(1) Parallel to the existing extended character set for Roman script (ISO/646)[9] machine-readable codes should be elaborated for those non-Roman scripts that produce the bulk of the world's present production of written records. Where such national codes already exist (e.g. for Russian Cyrillic script[10]) they should be adopted or, if necessary, amended and augmented to form the basis for an international exchange code.[11] These codes should be used solely for the transformation of the graphemes of a non-Roman script into one or more Roman letters and/or other signs that will represent them unambiguously and reversibly. In order to minimize the ambiguity inherent in the conventional Romanization of Chinese characters these should be represented by the numbers assigned to them in standard Chinese-English and Japanese-English dictionaries respectively.

(2) All existing ISO transliteration standards should be revised with the following aims in mind:
   (a) bringing them as nearly as possible into conformity with authorized transliteration schemes used in those countries in which a non-Roman script is the official one;[12]
   (b) making them applicable primarily for bibliographic purposes;
   (c) making them typographically as simple as possible, i.e. with minimal use of diacritical marks, and eliminating all unusual or technically difficult typographical provisions such as superscript or subscript letters or specially devised diacritical marks.

(3) Further ISO transliteration standards should be elaborated for scripts not yet covered (or still in the draft stage), especially, Japanese, Chinese, Devanagari (and related scripts), Korean, Thai, Burmese, Armenian, and Amharic, in decreasing order of importance.

(4) All international information retrieval systems and services should be urged to use these revised ISO standards without any modifications or exceptions. If this is found to be not practically feasible (e.g. because of technical limitations of the printing system) the system used in English-speaking countries should be used because (a) English is presently a kind of world-wide medium of communication, and (b) an English-oriented system will have no or few diacritical marks.

(5) All national organizations concerned with bibliographic control should be urged to use the international switching code of conversion but not to employ it for direct printout in their published retrieval tools. In bibliographies and catalogs, only the nationally accepted conversion scheme scheme should be used, expressing not only written characters but also their phonemic realization where necessary.

(6) All national bibliographies should be asked to provide entries in their respective vernaculars and scripts,[13] and in a form that would permit their reproduction and duplication by photomechanical methods. In addition, they should provide input to machine-readable switching tools, either directly in machine-readable form, or, if such equipment is not available locally, in eye-legible form to be converted to machine-readable codes by Western bibliographic agencies.

(7) Based on the input provided by national bibliographies, separate machine-readable data bases for bibliographic data in non-Roman scripts, converted to the international switching code, should be developed at focal points, from which they would be disseminated to libraries and bibliographic services. This task should probably be divided among the world's largest national libraries and bibliographic centers as well as among the ISDS centers, each of which would be responsible for the creation and maintenance of data bases for one or more non-Roman scripts.

These proposals recognize that

(a) a sizeable proportion of the world's written records are in non-Roman scripts;

(b) the major bibliographic services are concentrated in the Western world and use Roman script, yet they have a vital and legitimate interest in the coverage of all documents, irrespective of script;

(c) bibliographic control of documents in non-Roman scripts must be intelligible to users in the West, whether or not they are able to read other scripts and know languages other than their own; those users include librarians and bibliographers concerned less with the contents of documents than with their proper identification for purposes of acquisition, storage, retrieval, and dissemination;

(d) bibliographic control of documents in non-Roman scripts must, however, at the same time also make it possible to directly identify all documents in their original script and language;

(e) the goals set in paragraphs (c) and (d) above are not incompatible with each other, and can be achieved simultaneously by means of parallel listings of documents in separate sequences by script, and through the use of machine-readable switching codes.

TABLE 1

BOOK PRODUCTION BY SCRIPT 1973[1]

Rank	Script	Titles: Roman	Titles: Non-Roman	All titles: Percentage N = 566,000	All titles: Cumulative percentage	Non-Roman titles: Percentage N = 162,000	Non-Roman titles: Cumulative percentage
1	Roman	404,000		71.3	71.3		
2	Cyrillic		81,000	14.3	85.6	50.0	50.0
3	Japanese		35,500	6.3	91.9	21.9	71.9
4	Chinese[2]		11,200	1.9	93.8	6.9	78.8
5	Devanagari[3]		8,800	1.6	95.4	5.4	84.2
6	Arabic		7,500	1.3	96.7	4.6	88.8
7	Korean		7,400	1.3	98.0	4.6	93.4
8	Greek		2,300	0.4	98.4	1.4	94.8
9	Thai		2,200	0.4	98.8	1.4	96.2
10	Hebrew		2,000	0.4	99.2	1.2	97.4
11	Georgian		1,600	0.3	99.5	1.0	98.4
12	Burmese		1,500	0.3	99.8	0.9	99.3
13	Armenian		800	0.1	99.9	0.5	99.8
14	Amharic[4]		200	.	100.0	0.1	99.9
	Subtotal	404,000	162,000				
	Total		566,000				

(all rounded to nearest 100)

[1] Figures are based on the data for book production in 1973, published in Unesco Statistical Yearbook 1974 (the latest with a detailed breakdown by language), except as noted below.

[2] Data computed on the basis of an official report quoted in Library of Congress Information Bulletin, 34, no. 24, May 16, 1975, p. 203.

[3] Including 11 other Indian scripts closely related to or derived from Devanagari.

[4] Data for 1974. Source: Ethiopian Publications, 1975.

REFERENCES

1) Data for the number of periodicals are based on figures published by the British Library Lending Division. The 1962 estimates and the figure of 70 articles per year are from C.P. Bourne, "The world's technical journal literature: an estimate of volume, origin, language, field, indexing and abstracting." American Documentation 13 (1962):159-168. Derek J. de Solla Price and Suha Gursey in their paper "Some statistical results for the number of authors in the States of the United States and the nations of the world", published in Who is publishing in science?, 1977, p. 26-34, found that the relative number of publishing scholars rises much faster in the developing countries than in developing ones; many of these publish in their vernacular language and script, so that the relative number of articles in non-Roman scripts is probably rising steadily (this latter assumption is my own, not de Solla Price's).

2) For example, about one third of Thailand's annual production of books consists of collections of eulogies for deceased persons, printed on the occasion of their cremation ceremonies, and intended for relatives and friends. (Conference on Universal Bibliographic Control in Southeast Asia, Singapore, 1975. Papers and Proceedings. Singapore: Library Association of Singapore, 1975. p. 107)

3) Weinberg, B. "Transliteration in documentation." Journal of Documentation 30 (1974):18-31. - Spalding, C.S. "Romanization reexamined." Library Resources & Technical Services 21 (1977):3-12.

4) Wellisch, H.H. The conversion of scripts. New York: Wiley, 1978. See especially chapters 5 and 6. See also Wellisch, H.H. "Script conversion and bibliographic control of documents in dissimilar scripts: problems and alternatives." International Library Review 10 (1978):3-22.

5) Wellisch, H.H. "Script conversion practices in the world's libraries." International Library Review 8 (1976):55-84.

6) See ref. (2) and Wellisch, H.H. "Multiscript and multilingual bibliographic control: alternatives to Romanization." Library Resources and Technical Services 22 (1978) (In press).

7) Malinconico, S., et al. "Vernacular scripts in the NYPL automated bibliographic control system." Journal of Library Automation 10 (1977):205-225.

8) The few instances in which pages are numbered by a system other than Arabic digits can for all practical purposes be disregarded, since it is always possible to render them in Arabic digits.

9) ISO/R646-1973. 7-bit coded character set for information processing interchange.

10) [USSR standard for 7-bit coded character sets. GOST 13052-67.]

11) The standard for Cyrillic script (see ref. 10), for example, would have to be extended to comprise also letters used in Bulgarian, Serbian, and Macedonian.

12) This would mean the endorsement of the Pin-yin system for Chinese, and the Kunrei system for Japanese -- not because they are necessarily better than the Wade-Giles or Hepburn systems, but because they represent the intentions of the governments of the peoples concerned as opposed to what Westerners think ought to be done.

13) This is in conformity with resolutions adopted at the Conference on Universal Bibliographic Control in Southeast Asia, Singapore, 1975 (see ref. (2)), and at the International Congress on National Bibliographies in Paris, 1977, both of which emphasized the need for records in the original language and script of a nation, accompanied by a Romanization.